(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 November 2000 (30.11.2000)

PCT

(10) International Publication Number
WO 00/72258 A2

(51) International Patent Classification[7]: G06T

(21) International Application Number: PCT/US00/14246

(22) International Filing Date: 24 May 2000 (24.05.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/135,481 24 May 1999 (24.05.1999) US
60/140,061 21 June 1999 (21.06.1999) US

(71) Applicant *(for all designated States except US)*: CELLOMICS, INC. [US/US]; 635 William Pitt Way, Pittsburgh, PA 15328 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: GOUGH, Albert, H. [US/US]; 2014 Louise Drive, Glenshaw, PA 15116 (US). LAPETS, Oleg P. [RU/US]; 1937 Shady Oak Circle, Allison Park, PA 15101 (US). BRIGHT, Gary [US/US]; 586 Sturbridge Drive, Highland Heights, OH 44143 (US).

(74) Agent: HARPER, David, S.; McDonnell Boehnen Hulbert & Berghoff, Suite 3200, 300 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/72258 A2

(54) Title: METHOD AND SYSTEM FOR GENERAL PURPOSE ANALYSIS OF EXPERIMENTAL DATA

(57) Abstract: Methods and system for general purpose analysis of images acquired from experimental data collected with automated feature-rich, high-throughput experimental data collection systems. A set of pre-determined general assay features is presented. An assay feature includes one or more measurements for an object in a digital photographic image acquired from the experimental data. The set of pre-determined general assay features includes object features, aggregate features and general purpose image processing features. A set of desired assay features is selected from the set of features. A set of images is processed using the desired assay features from the selected set of assay features. The methods and system help provide a general purpose assay development tool. The methods and system allow a biologist, other scientist or lab technician not trained in image processing techniques to quickly and easily design protocols and assays to analyze images acquired from experimental data (e.g., cells). The methods and system may improve the identification, selection, validation and screening of new drug compounds that have been applied to populations of cells. The methods and system may also be used to provide new bioinformatic techniques to manipulate experimental data including multiple digital photographic images.

METHOD AND SYSTEM FOR GENERAL PURPOSE ANALYSIS OF EXPERIMENTAL DATA

CROSS REFERENCES TO RELATED APPLICATIONS

This applications claims priority from U.S. Provisional Applications No. 60/135,481, filed on May 24, 1999, and 60/140,061, filed on June 21, 1999.

COPYRIGHT AUTHORIZATION

FIELD OF THE INVENTION

This invention relates to analyzing experimental data. More specifically, it relates to methods and system for general purpose analysis of images from experimental data collected with automated feature-rich, high-throughput experimental data collection systems.

BACKGROUND OF THE INVENTION

Historically, the discovery and development of new drugs has been an expensive, time consuming and inefficient process. With estimated costs of bringing a single drug to market requiring an investment of approximately 8 to 12 years and approximately $350 to $500 million, the pharmaceutical research and development market is in need of new technologies that can streamline the drug discovery process. Companies in the pharmaceutical research and development market are under fierce

pressure to shorten research and development cycles for developing new drugs, while at the same time, novel drug discovery screening instrumentation technologies are being deployed, producing a huge amount of experimental data.

Innovations in automated screening systems for biological and other research are capable of generating enormous amounts of data. The massive volumes of feature-rich data being generated by these systems and the effective management and use of information from the data has created a number of very challenging problems. As is known in the art, "feature-rich" data includes data wherein one or more individual features of an object of interest (e.g., a cell) can be collected. To fully exploit the potential of data from high-volume data generating screening instrumentation, there is a need for new informatic and bioinformatic tools.

Identification, selection, and validation of targets for the screening of new drug compounds is often completed at a nucleotide level using sequences of Deoxyribonucleic Acid ("DNA"), Ribonucleic Acid ("RNA") or other nucleotides. "Genes" are regions of DNA, and "proteins" are the products of genes. The existence and concentration of protein molecules typically helps determine if a gene is "expressed" or "repressed" in a given situation. Responses to natural and artificial compounds as indicated by changes in gene expression are typically used to improve existing drugs, and develop new drugs. Changes in binding between proteins are also used to screen compounds for biological activity. However, it is often more appropriate to determine the effect of a new compound on a cellular level instead of a nucleotide or protein level.

Cells are the basic units of life and integrate information from DNA, RNA, proteins, metabolites, ions and other cellular components. New compounds that may look promising at a nucleotide or protein level may be toxic at a cellular or organism

level. Florescence-based reagents can be applied to cells to determine ion concentrations, membrane potentials, enzyme activities, gene expression, as well as the presence of metabolites, proteins, lipids, carbohydrates, and other cellular components.

There are two types of cell screening methods that are typically used: (1) fixed cell screening; and (2) live cell screening. For fixed cell screening, initially living cells are treated with experimental compounds being tested. After application of a desired compound the cells are incubated for a given time and then "fixed" to preserve a final cell state for later analysis. Live cell screening usually requires environmental control of the cells (e.g., temperature, humidity, gases, etc.) since before, during and after application of a desired compound, the cells are kept in a controlled environment until data collection is complete. Fixed cell assays allow spatial measurements to be acquired, but only at one point in time. Live cell assays allow both spatial and temporal measurements to be acquired.

As is known in the art, a "cell assay" is a specific implementation of image processing methods used to analyze images of cells and return results related to the biological processes being examined. As is known in the art, a "cell protocol" specifies a series of system settings including a type of analysis instrument, a cell assay, dyes used to measure biological markers in cells, cell identification parameters and other general image processing parameters used to collect cell data.

The spatial and temporal frequency of chemical and molecular information present within cells makes it possible to extract feature-rich cell information from populations of cells. For example, multiple molecular and biochemical interactions, cell kinetics, changes in sub-cellular distributions, changes in cellular morphology, changes in individual cell subtypes in mixed populations, changes and sub-cellular

molecular activity, changes in cell communication, and other types of cell information can be acquired.

The types of biochemical and molecular cell-based assays now accessible through fluorescence-based reagents is expanding rapidly. The need for automatically extracting additional information from a growing list of cell-based assays has allowed automated platforms for feature-rich assay screening of cells to be developed. For example, the ArrayScan System by Cellomics, Inc. of Pittsburgh, Pennsylvania, is one such feature-rich cell screening system. Cell based systems such as FLIPR, by Molecular Devices, Inc. of Sunnyvale, California, FMAT, of PE Biosystems of Foster City, California, ViewLux by EG&G Wallac, now a subsidiary of Perkin-Elmer Life Sciences of Gaithersburg, Maryland, and others also generate large amounts of data and photographic images that would benefit from efficient data management solutions. Photographic images are typically collected using a digital camera, but can also be generated by scanning systems such as confocal light microscope systems. A single photographic image may take up as much as 512 Kilobytes ("KB") or more of storage space as is explained below. Collecting and storing a large number of photographic images ads to the data problems encountered when using high throughput systems. For more information on fluorescence based systems, see "Bright ideas for high-throughput screening – One-step fluorescence HTS assays are getting faster, cheaper, smaller and more sensitive," by Randy Wedin, Modern Drug Discovery, Vol. 2(3), pp. 61-71, May/June 1999.

Such automated feature-rich cell screening systems and other systems known in the art typically include microplate scanning hardware, fluorescence excitation of cells, fluorescence emission optics, a microscope with a camera, data collection, data storage and data display capabilities. For more information on feature-rich cell

screening see "High content fluorescence-based screening," by Kenneth A. Guiliano, et al., Journal of Biomolecular Screening, Vol. 2, No. 4, pp. 249-259, Winter 1997, ISSN 1087-0571, "PTH receptor internalization," Bruce R. Conway, et al., Journal of Biomolecular Screening, Vol. 4, No. 2, pp. 75-68, April 1999, ISSN 1087-0571, "Fluorescent-protein biosensors: new tools for drug discovery," Kenneth A. Giuliano and D. Lansing Taylor, Trends in Biotechnology, ("TIBTECH"), Vol. 16, No. 3, pp. 99-146, March 1998, ISSN 0167-7799, all of which are incorporated by reference.

An automated feature-rich cell screening system typically automatically scans a microplate with multiple wells and acquires multi-color fluorescence data of cells at one or more instances of time at a pre-determined spatial resolution. Automated feature-rich cell screening systems typically support multiple channels of fluorescence to collect multi-color fluorescence data and may also provide the ability to collect cell feature information on a cell-by-cell basis including such features as the brightness, size and shape of cells and sub-cellar measurements of organelles within a cell.

The collection of data from high throughput screening systems typically produces a very large quantity of data and presents a number of bioinformatics problems. As is known in the art, "bioinformatic" techniques are used to address problems related to the collection, processing, storage, retrieval and analysis of biological information including cellular information. Bioinformatics is defined as the systematic development and application of information technologies and data processing techniques for collecting, analyzing and displaying data acquired by experiments, modeling, database searching, and instrumentation to make observations about biological processes.

The need for efficient data management is not limited to feature-rich cell screening systems or to cell based arrays. Virtually any instrument that runs High

Throughput Screening ("HTS") assays also generate large amounts of data. For example, with the growing use of other data collection techniques such as DNA arrays, bio-chips, microscopy, micro-arrays, gel analysis, the amount of data collected, including photographic image data is also growing exponentially. As is known in the art, a "bio-chip" is a stratum with hundreds or thousands of absorbent micro-wells on its surface. A micro-well includes a specific point of attachment that may or may not have any depth. A single bio-chip may contain 10,000 or more micro-gels. When performing an assay test, each micro-well on a bio-chip is like a micro-test tube or a well in a microplate. A bio-chip provides a medium for analyzing known and unknown biological (e.g., nucleotides, cells, etc.) samples in an automated, high-throughput screening system.

Although a wide variety of data collection techniques can be used, cell-based high throughput screening systems are used as an example to illustrate some of the associated data management problems encountered by virtually all high throughput screening systems. Collecting feature-rich cell data from a microplate plate used for feature-rich screening typically includes 96 to 1536 individual wells. As is known in the art, a "microplate" is a flat, shallow dish that stores multiple samples for analysis. A "well" is a small area in a microplate used to contain an individual sample for analysis. Each well may be divided into multiple fields. A "field" is a sub-region of a well that represents a field of vision (i.e., a zoom level) for a photographic microscope. Each well is typically divided into one to sixteen fields, or more

Each field typically will have between one and six photographic images taken of it, each using a different light filter to capture a different wavelength of light for a different fluorescence response for desired cell components. In each field, a pre-determined number of cells are selected to analyze. The number of cells will vary

(e.g., between one and one hundred or more). For each cell, multiple cell features are collected. The cell features may include features such as size, shape, brightness, pattern, etc. of a cell.

There are a number of problems associated with analyzing experimental data collected from feature-rich cell screening systems. One problem is that a biologist may desire to create his/her own cell assay to analyze biological processes associated with cells. However, most biologist do not have the expertise required to implement image processing methods necessary to complete his/her cell assay.

Another problem is that a biologist may desire to develop two or more different cell assays run at the same time to focus on different cell information. For example, for a first cell assay it may be necessary to collect cell feature data including cell shape, cell size and cell diameter data for a desired experiment by analyzing cell image data. For a second cell assay, it may be desirable to collect skewness and kurtosis for a desired cell feature by analyzing cell image data. However, analysis tools known in the art do not allow a biologist to select his/her own image processing techniques to create a cell assay outside of a fixed list of image processing techniques available with the analysis tool. That is, a biologist may desire to analyze skewness and kurtosis, but his/her analysis tool may only provide image processing techniques for analyzing cell shape, and cell size.

Another problem is that many image processing tools can not be easily interfaced with existing feature-rich cell screening systems. Many image processing tools known in the art are proprietary and are not adaptable for general use with existing feature-rich cell screening systems. This also limits the ability of a biologist to create a cell assay for a desired experiment.

Another problem is that even if image processing packages known in the art are used, a biologist or other scientist, has to select not only image processing routines to accomplish an assay feature measurement, but also choose from a large number of image processing options for the image processing routines. This may create additional confusion or frustration on the part of the biologist as the biologist may not know what image processing options are the most appropriate for a give assay feature.

Thus, it is desirable to provide a general purpose analysis tool that allows virtually any cell assay to be created by a biologist. The general purpose tool should provide image processing techniques for a cell assay created by a biologist, without requiring the biologist, other scientist or analyst have any in-depth knowledge of image processing techniques.

SUMMARY OF THE INVENTION

In accordance with preferred embodiments of the present invention, some of the problems associated with analyzing image acquired from feature-rich experimental data are overcome. Methods and system for general purpose analysis of images acquired from experimental data are presented.

One aspect of the invention includes a method for presenting assay features associated with a pre-determined set of image processing routines for analyzing experimental data including images. The pre-determined set of image processing routines includes only a limited set of options available for processing an image. Another aspect of the invention includes a method for analyzing experimental data including images using a set of selected assay features selected from a set of pre-determined assay features to help analyze image data. The set of selected assay features are processed in a pre-determined order appropriate for analysis of image data.

A pre-determined set of general assay features is presented. An assay feature includes one or more measurements for an object in a digital photographic image acquired from the experimental data. The set of general assay features includes object features, aggregate features and general purpose image processing features. A set of desired assay features is selected from the pre-determined set of general assay features. A set of images is processed using the desired assay features from the selected set of general assay features. Such general assay features (e.g., length, width, height, etc.) are common image processing features that are useful for virtually any assay or protocol that may be developed to obtain measurements from experimental data. The general assay features presented typically include only a few of the many possible image processing options that could be used to take such measurements from

a digital image, thereby helping to reduce confusion associated selecting such image processing options.

The methods and system may help provide a general purpose assay development tool. The methods and system may allow a biologist, other scientist or lab technician not trained in image processing techniques to quickly and easily design protocols and assays to analyze images acquired from experimental data (e.g., cells). The methods and system may improve the identification, selection, validation and screening of new experimental compounds (e.g., drug compounds). The methods and system may also be used to provide new bioinformatic techniques used to make observations about experimental data including multiple digital photographic images.

The foregoing and other features and advantages of preferred embodiments of the present invention will be more readily apparent from the following detailed description. The detailed description proceeds with references to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the present invention are described with reference to the following drawings, wherein:

FIG. 1A is a block diagram illustrating an exemplary experimental data storage system;

FIG. 1B is a block diagram illustrating an exemplary experimental data storage system;

FIG. 2 is a block diagram illustrating an exemplary array scan module architecture;

FIG. 3 is a flow diagram illustrating a method for selecting assay features for experimental data.

FIG. 4 is a flow diagram illustrating a method for selecting assay features for images acquired from experimental data;

FIG. 5 is a block diagram illustrating an exemplary graphical user interface for selecting object features;

FIG. 6 is a block diagram illustrating an exemplary graphical user interface for selecting general image processing operations; and

FIG. 7 is a block diagram illustrating a screen display for graphically displaying images processed using a desired set of assay features.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Exemplary data storage system

FIG. 1A illustrates an exemplary data storage system 10 for preferred embodiments of the present invention. The exemplary data storage system 10 includes an analysis instrument 12, connected to a client computer 18, a shared database 24 and a data store archive 30 with a computer network 40. The analysis instrument 12 includes any scanning instrument capable of collecting feature-rich experimental data, such as nucleotide, protein, cell or other experimental data, or any analysis instrument capable of analyzing feature-rich experimental data. As is known in the art, "feature-rich" data includes data wherein one or more individual features of an object of interest (e.g., a cell) can be collected. The client computer 18 is any conventional computer including a display application that is used to lead a scientist or lab technician through data analysis. The shared database 24 is a multi-user, multi-view relational database that stores data from the analysis instrument 12. The data archive 30 is used to provide virtually unlimited amounts of "virtual" disk space with a multi-layer hierarchical storage management system. The computer network 40 is any fast Local Area Network ("LAN") (e.g., capable of data rates of 100 Mega-bit per second or faster). However, the present invention is not limited to this embodiment and more or fewer, and equivalent types of components can also be used. Data storage system 10 can be used for virtually any system capable of collecting and/or analyzing feature-rich experimental data from biological and non-biological experiments.

FIG. 1B illustrates an exemplary data storage system 10' for one preferred embodiment of the present invention with specific components. However, the present invention is not limited to this one preferred embodiment, and more or fewer, and equivalent types of components can also be used. The data storage system 10' includes

one or more analysis instruments 12, 14, 16, for collecting and/or analyzing feature-rich experimental data, one or more data client computers, 18, 20, 22, a shared database 24, a data store server 26, and a shared database file server 28. A data store archive 30 includes any of a disk archive 32, an optical jukebox 34 or a tape drive 36. The data store archive 30 can be used to provide virtually unlimited amounts of "virtual" disk space with a multi-layer hierarchical storage management system without changing the design of any databases used to stored collected experimental data as is explained below. The data store archive 30 can be managed by an optional data archive server 38. Data storage system 10' components are connected by a computer network 40. However, more or fewer data store components can also be used and the present invention is not limited to the data storage system 10' components illustrated in FIG. 1B.

In one exemplary preferred embodiment of the present invention, data storage system 10' includes the following specific components. However, the present invention is not limited to these specific components and other similar or equivalent components may also be used. Analysis instruments 12, 14, 16, comprise a feature-rich array scanning system capable of collecting and/or analyzing experimental data such as cell experimental data from microplates, DNA arrays or other chip-based or bio-chip based arrays. Bio-chips include any of those provided by Motorola Corporation of Schaumburg, Illinois, Packard Instrument, a subsidiary of Packard BioScience Co. of Meriden, Connecticut, Genometrix, Inc. of Woodlands, Texas, and others.

Analysis instruments 12, 14, 16 include any of those provided by Cellomics, Inc. of Pittsburgh, Pennsylvania, Aurora Biosciences Corporation of San Diego, California, Molecular Devices, Inc. of Sunnyvale, California, PE Biosystems of Foster City, California, Perkin-Elmer Life Sciences of Gaithersburg, Maryland, and others. The one or more data client computers, 18, 20, 22, are conventional personal

computers that include a display application that provides a Graphical User Interface ("GUI") to a local hard disk, the shared database 24, the data store server 26 and/or the data store archive 30. The GUI display application is used to lead a scientist or lab technician through standard analyses, and supports custom and query viewing capabilities. The display application GUI also supports data exported into standard desktop tools such as spreadsheets, graphics packages, and word processors.

The data client computers 18, 20, 22 connect to the store server 26 through an Open Data Base Connectivity ("ODBC") connection over network 40. In one embodiment of the present invention, computer network 40 is a 100 Mega-bit ("Mbit") per second or faster Ethernet, Local Area Network ("LAN"). However, other types of LANs could also be used (e.g., optical or coaxial cable networks). In addition, the present invention is not limited to these specific components and other similar components may also be used.

As is known in the art, OBDC is an interface providing a common language for applications to gain access to databases on a computer network. The store server 26 controls the storage based routines plus an underlying Database Management System ("DBMS").

The shared database 24 is a multi-user, multi-view relational database that stores summary data from the one or more analysis instruments 12, 14, 16. The shared database 24 uses standard relational database tools and structures. The data store archive 30 is a library of image and feature database files. The data store archive 30 uses Hierarchical Storage Management ("HSM") techniques to automatically manage disk space of analysis instruments 12, 14, 16 and the provide a multi-layer hierarchical storage management system. For more information on data storage system 10 and 10' see, co-pending application number 09/437,976, entitled

"Methods and System for Efficient Collection and Storage of Experimental Data," assigned to the same Assignee as the present invention, and incorporated herein by reference.

An operating environment for components of the data storage system 10 and 10' for preferred embodiments of the present invention include a processing system with one or more high-speed Central Processing Unit(s) ("CPU") and a memory. In accordance with the practices of persons skilled in the art of computer programming, the present invention is described below with reference to acts and symbolic representations of operations or instructions that are performed by the processing system, unless indicated otherwise. Such acts and operations or instructions are referred to as being "computer-executed" or "CPU executed."

It will be appreciated that acts and symbolically represented operations or instructions include the manipulation of electrical signals by the CPU. An electrical system represents data bits which cause a resulting transformation or reduction of the electrical signals, and the maintenance of data bits at memory locations in a memory system to thereby reconfigure or otherwise alter the CPU's operation, as well as other processing of signals. The memory locations where data bits are maintained are physical locations that have particular electrical, magnetic, optical, or organic properties corresponding to the data bits.

The data bits may also be maintained on a computer readable medium including magnetic disks, optical disks, organic memory, and any other volatile (e.g., Random Access Memory ("RAM")) or non-volatile (e.g., Read-Only Memory ("ROM")) mass storage system readable by the CPU. The computer readable medium includes cooperating or interconnected computer readable medium, which exist exclusively on the processing system or be distributed among multiple

interconnected processing systems that may be local or remote to the processing system.

Array scan module architecture

FIG. 2 is a block diagram illustrating an exemplary array scan module 42 architecture. The array scan module 42, such as one associated with analysis instrument 12, 14, 16 (FIG. 1B) includes software/hardware that is divided into four functional groups or modules. However, more of fewer functional modules can also be used and the present invention is not limited to four functional modules. The Acquisition Module 44 controls a robotic microscope and digital camera, acquires images and sends the images to the Assay Module 46. The Assay Module 46 "reads" the images, creates graphic overlays, interprets the images collects feature data and returns the new images and feature data extracted from the images back to the Acquisition Module 44. The Acquisition Module 44 passes the image and interpreted feature data to the Data Base Storage Module 48. The Data Base Storage Module 48 saves the image and feature information in a combination of image files and relational database records. The client computers 18, 20, 22 use the Data Base Storage Module 48 to access feature data and images for presentation and data analysis by the Presentation Module 50. The Presentation Module 50 includes a display application with a GUI as was discussed above.

Selecting features for images acquired from experimental data

FIG. 3 is a flow diagram illustrating a Method 52 for selecting assay features for experimental data. In FIG. 3 at Step 54, multiple pre-determined assay features for analyzing images acquired from experimental data are presented. An assay feature includes one or more measurements for an object in an image acquired from the experimental data. At Step 56, a set of desired assay features selected from the

multiple presented assay features are received. At Step 58, one or more image processing routines from a library of image processing routines are selected for an assay feature from the set of desired assay features. The one or more image processing routines are used to accomplish the selected assay feature. At Step 60, the one or more image processing routines are associated with the assay feature. At Step 62, a loop is entered to repeat steps 58 and 60 for assay features in the set of selected assay features.

Method 52 is illustrated with one specific embodiment of the present invention. However, the present invention is not limited to such an embodiment and other embodiments can also be used.

In such an embodiment, at Step 54 multiple pre-determined assay features for analyzing digital photographic images (hereinafter "images") acquired from experimental data for an assay are presented by analysis instruments 12, 14, 16 (FIG. 1B) or by client computers 18, 20, 22 (FIG. 1B). In one embodiment of the present invention, the multiple pre-determined assay features include object features (See, e.g., FIG. 5). An "object" feature operates on an individual object (e.g., a cell) or an object component (e.g., cell membrane, cell nucleus, etc.) In another embodiment of the present invention, the multiple pre-determined assay features include object features and aggregate features. An "aggregate" feature includes assay features that operate on multiple objects (e.g., number of objects, average value of a feature, standard deviation value of a feature, etc.). In another embodiment of the present invention, the multiple pre-determined assay features include only aggregate features.

In one specific embodiment of the present invention, the multiple pre-determined assay features presented at Step 54 include general assay features that can be used by virtually any biologist, other scientist or analyst to analyze measurements

from objects (e.g., cells) in images collected from experimental data. Such general assay features (e.g., length, width, height, etc.) are common image processing features that are useful for virtually any assay or protocol that may be developed to obtain measurements from experimental data. In such an embodiment, the general assay features presented typically include only a few of the many possible image processing options that could be used to take measurements from a digital image.

For example, an assay feature for a simple measurement such as determining an object's length, may include multiple different types of image processing thresholds (e.g., a number of pixels, types of pixels, type of object components in/around a desired object, etc. to be included for the object to determine its length). In one embodiment of the present invention, two image processing thresholds (e.g., a minimum and a maximum) value may be presented to a user for determining an object's length. Other image processing thresholds are handled internally without presenting such information to a user.

The general assay features and limited image processing options for the general assay features presented allow a biologist, other scientist or analyst without much image processing experience to easily and quickly create assays and protocols. Since general assay features and limited image processing options are presented, instead of specific assay features with many different options, a user with limited image processing experience is less likely to get confused when he/she is creating an assay or protocol.

In one specific embodiment of the present invention, the general assay features associated with image processing options are presented in a specific ordering. However, the present invention is not limited to such an embodiment with such a

specific ordering. This specific ordering may also help a user with limited knowledge of image processing select the appropriate options for a desired assay or protocol.

Typically an assay will include two or more channels. A "channel" is a specific configuration of optical filters and channel specific parameters that are used to acquire an image. In a typical assay, different fluorescent dyes are used to label different cell structures. The fluorescent dyes emit light at different wavelengths. Channels are used to acquire photographic images for different dye emission wavelengths.

Given a digitized image including one or more objects (e.g., cells), there are typically two phases to analyzing an image and extracting feature data as feature measurements. The first phase is typically called "image segmentation" or "object isolation," in which a desired object is isolated from the rest of the image. The second phase is typically called "feature extraction," wherein measurements of the objects are calculated. A feature is typically a function of one or more measurements, calculated so that it quantifies a significant characteristic of an object. Typical object measurements include size, shape, intensity, texture, location, and others.

For each measurement, several features are commonly used to reflect the measurement. The "size" of an object can be represented by its area, perimeter, boundary definition, length, width, etc. The "shape" of an object can be represented by its rectangularity (e.g., length and width aspect ratio), circularity (e.g., perimeter squared divided by area, bounding box, etc.), moment of inertia, differential chain code, Fourier descriptors, etc. The "intensity" of an object can be represented by a summed average, maximum or minimum grey levels of pixels in an object, etc. The "texture" of an object quantifies a characteristic of grey-level variation within an object and can be represented by statistical features including standard deviation,

variance, skewness, kurtosis and by spectral and structural features, etc. The "location" of an object can be represented by an object's center of mass, horizontal and vertical extents, etc. with respect to a pre-determined grid system. For more information on digital image feature measurements, see: "Digital Image Processing," by Kenneth R. Castleman, Prentice-Hall, 1996, ISBN-0132114674, "Digital Image Processing: Principles and Applications," by G. A. Baxes, Wiley, 1994, ISBN-0471009490 , "Digital Image Processing," by William K. Pratt, Wiley and Sons, 1991, ISBN-0471857661, or "The Image Processing Handbook – 2nd Edition," by John C. Russ, CRC Press, 1991, ISBN-0849325161, the contents of all of which are incorporated by reference.

In one exemplary preferred embodiment of the present invention, Method 52 is used to analyze cell image data and cell feature data from "wells" in a "microplate." In another preferred embodiment of the present invention, Method 52 is used to analyze cell image and cell feature data from micro-gels in a bio-chip. As is known in the art, a "microplate" is a flat, shallow dish that stores multiple samples for analysis and typically includes 96 to 1536 individual wells. A "well" is a small area in a microplate used to contain an individual sample for analysis.

Each well may be divided into multiple fields. A "field" is a sub-region of a well that represents a field of vision (i.e., a zoom level) for a photographic microscope. Each well is typically divided into one to sixteen fields, or more. Each field typically will have between one and six photographic images taken of it, each using a different light filter to capture a different wavelength of light for a different fluorescence response for desired cell components. However, the present invention is not limited to such an embodiment, and other containers (e.g., varieties of biological chips, such as DNA chips, micro-arrays, and other containers with multiple sub-

containers), sub-containers can also be used to collect image data and feature data from other than cells.

In one embodiment of the present invention, Step 54 includes presenting a set of static assay features in a uniform manner on a graphical user interface for every user. In such an embodiment, the set of static assay features cannot be modified by a user. In another embodiment of the present invention, Step 54 is optionally split into two sub-steps. In a first sub-step, a user first selects a desired set of assay feature names from a list of assay features. In a second sub-step the desired set of assay feature names is dynamically presented on graphical user interface specifically for the user. In such an embodiment, a user can dynamically modify the set of assay features that will actually be presented and used instead of receiving a set of static assay features that cannot be modified by a user. Any assay features selected by a user from a list of assay features are also associated with one or more image processing routines as is described for Step 58 below.

As was described above, an assay feature includes one or more measurements for an object in an image acquired from experimental data. In one exemplary embodiment of the present invention, objects in the images acquired from experimental data include, but are not limited to, cells. Exemplary object features for cells are illustrated in Table 1. However, other object features and can also be used and the present invention is not limited to the cell features illustrated in Table 1. Virtually any object feature can be presented at Step 54.

CELL SIZE
CELL SHAPE
CELL INTENSITY
CELL TEXTURE
CELL LOCATION
CELL AREA

CELL PERIMETER
CELL SHAPE FACTOR
CELL EQUIVALENT DIAMETER
CELL LENGTH
CELL WIDTH
CELL INTEGRATED FLUORESCENCE INTENSITY
CELL MEAN FLUORESCENCE INTENSITY
CELL VARIANCE
CELL SKEWNESS
CELL KURTOSIS
CELL MINIMUM FLUORESCENCE INTENSITY
CELL MAXIMUM FLUORESCENCE INTENSITY
CELL GEOMETRIC CENTER
CELL X-COORDINATE OF A GEOMETRIC CENTER
CELL Y-COORDINATE OF A GEOMETRIC CENTER

Table 1.

Step 54 also includes presenting aggregate features. Aggregate features are features associated with a collection of objects such as a population of cells. In one exemplary embodiment of the present invention, the aggregate features include, but are not limited to, any of the well summary data for a microplate including cells illustrated in Table 2. However, the present invention is not limited to presenting aggregate features for the well summary data illustrated in Table 2. Virtually any summary data for aggregate features can be presented. In Table 2, a "SPOT" indicates a small region of fluorescent response intensity as a measure of biological activity.

WELL CELL SIZES
WELL CELL SHAPES
WELL CELL INTENSITIES
WELL CELL TEXTURES
WELL CELL LOCATIONS
WELL NUCLEUS AREA
WELL SPOT COUNT
WELL AGGREGATE SPOT AREA
WELL AVERAGE SPOT AREA
WELL MINIMUM SPOT AREA
WELL MAXIMUM SPOT AREA
WELL AGGREGATE SPOT INTENSITY
WELL AVERAGE SPOT INTENSITY
WELL MINIMUM SPOT INTENSITY
WELL MAXIMUM SPOT INTENSITY
WELL NORMALIZED AVERAGE SPOT INTENSITY
WELL NORMALIZED SPOT COUNT
WELL NUMBER OF NUCLEI

WELL NUCLEUS AGGREGATE INTENSITY
WELL DYE AREA
WELL DYE AGGREGATE INTENSITY
WELL NUCLEUS INTENSITY
WELL CYTOPLASM INTENSITY
WELL DIFFERENCE BETWEEN NUCLEUS AND CYTOPLASM INTENSITY
WELL NUCLEUS BOX-FILL RATIO
WELL NUCLEUS PERIMETER SQUARED AREA
WELL NUCLEUS HEIGHT/WIDTH RATIO
WELL CELL COUNT

Table 2.

The aggregate features can also include, but are not limited to, microplate summary data for cells illustrated in Table 3. In Table 3, "MEAN" indicates a statistical mean and "STDEV" indicates a statistical standard deviation, known in the art, and a "SPOT" indicates a small region of fluorescent response intensity as a measure of biological activity.

MEAN SIZE OF CELLS
MEAN SHAPES OF CELLS
MEAN INTENSITY OF CELLS
MEAN TEXTURE OF CELLS
LOCATION OF CELLS
NUMBER OF CELLS
NUMBER OF VALID FIELDS
STDEV NUCLEUS AREA
MEAN SPOT COUNT
STDEV SPOT COUNT
MEAN AGGREGATE SPOT AREA
STDEV AGGREGATE SPOT AREA
MEAN AVERAGE SPOT AREA
STDEV AVERAGE SPOT AREA
MEAN NUCLEUS AREA
MEAN NUCLEUS AGGREGATE INTENSITY
STDEV AGGREGATE NUCLEUS INTENSITY
MEAN DYE AREA
STDEV DYE AREA
MEAN DYE AGGREGATE INTENSITY
STDEV AGGREGATE DYE INTENSITY
MEAN MINIMUMSPOT AREA
STDEV MINIMUM SPOT AREA
MEAN MAXIMUM SPOT AREA
STDEV MAXIMUM SPOT AREA
MEAN AGGREGATE SPOT INTENSITY
STDEV AGGREGATE SPOT INTENSITY
MEAN AVERAGE SPOT INTENSITY
STDEV AVERAGE SPOT INTENSITY
MEAN MINIMUM SPOT INTENSITY
STDEV MINIMUM SPOT INTENSITY
MEAN MAXIMUM SPOT INTENSITY
STDEV MAXIMUM SPOT INTENSITY
MEAN NORMALIZED AVERAGE SPOT INTENSITY

STDEV NORMALIZED AVERAGE SPOT INTENSITY
MEAN NORMALIZED SPOT COUNT
STDEV NORMALIZED SPOT COUNT
MEAN NUMBER OF NUCLEI
STDEV NUMBER OF NUCLEI
NUCLEI INTENSITIES
CYTOPLASM INTENSITIES
DIFFERENCE BETWEEN NUCLEI AND CYTOPLASM INTENSITIES
NUCLEI BOX-FILL RATIOS
NUCLEI PERIMETER SQUARED AREAS
NUCLEI HEIGHT/WIDTH RATIOS
WELL CELL COUNTS

Table 3.

At Step 56, a set of assay features selected from the presented assay features are received on the analysis instruments 12, 14, 16 or client computers 18, 20, 22. For example, set of assay features selected from the multiple presented assay features may include object features for "cell perimeter," "cell width" and "cell length." (e.g., from Table 1).

At Step 58, one or more image processing routines from a library of image processing routines are selected for an assay feature from the set of selected assay features. The one or more image processing routines are used to accomplish the selected assay feature. To accomplish the "cell length" feature, one or more image processing routines are called from a library of image processing routines to accomplish the "cell length" feature. For example, image processing routines including "select_object( )," "object_boundingbox ( )," "object_rotate180 ( )," and "object_longest_side ( )" (e.g., see length feature in Table 6) may be selected from a library of image processing.

As is known in the art, there are many libraries of image processing routines. See for example, AnVisilog (Image Processing/Analysis Library), by NoesisVision, Inc. at the Universal Resource Locator ("URL") "www.noesisvision.com," MIL (Matrox Imaging Library) by Matrox Electronic Systems Ltd. At the URL "www.matrox.com," ImagePro (Image Processing/Analysis Library) and Optimas (

Image Processing/Analysis Library) by, MediaCybernetics, at the URL "www.mediacy.com," and others. Any of these image processing libraries or other known in the art can be used with the present invention.

At Step 60, the one or more image processing routines are associated with the selected feature. For example the "cell length" feature is associated with the image processing routines "select_object( )," "object_boundingbox ( )," "object_rotate180 ( )," and "object_longest_side ( )" (e.g., see length feature in Table 6).

At Step 62, a loop is entered to repeat steps 58 and 60 for assay features in the selected set of assay features. For example, after the cell length feature is associated with the image processing routines, the cell width and cell perimeter features are also associated with image processing routines by repeating steps 58 and 60.

Method 52 allows a biologist, other scientist or analyst not trained in image processing to create assays and protocols to analyze experimental data. Method 52 can be used to analyze images collected from feature-rich cell experimental data generated by HTS systems.

Processing selected assay features for images acquired from experimental data

FIG. 4 is a flow diagram illustrating a Method 64 for selecting assay features for images acquired from experimental data. At Step 66, a set of images is acquired from experimental data on an analysis device. At Step 68, a set of assay features is selected from a set of multiple presented assay features to analyze the set of images. An assay feature includes one or more measurements for an object in an image acquired from the experimental data. A presented assay feature is associated with one or more image processing routines from a library of image processing routines to accomplish the assay feature. At Step 70, processing of the set of images using the

selected set of assay features is requested. At Step 72, results are received from the processing of the set of images using the selected set of assay features.

Method 64 is illustrated with one specific embodiment of the present invention. However, the present invention is not limited to such an embodiment and other embodiments can also be used.

In such an embodiment, at Step 66, a set of images (e.g., for cells or components of cells acquired from cell experimental data) is acquired on analysis instruments 12, 14, 16 or client computers 18, 20, 22 (e.g., FIGS. 1A and 1B). In one embodiment of the present invention, there are two ways to acquire images: (1) from prepared samples; or (2) from stored image sets.

Images are acquired automatically from a feature rich array scanning system (e.g., using array scan module 42 of FIG. 2) as an experiment is being conducted. Images are acquired from stored images sets after a desired experiment has been run by a feature rich array scanning system and the results have been saved in a shared database 24 or a store archive 30, or local hard drive.

FIG. 5 is a block diagram illustrating an exemplary graphical user interface 74 presented on the analysis instruments 12, 14, 16 or client computers 18, 20, 22 for selecting object features at Step 68. The graphical user interface 74 includes graphical entities such as graphical check boxes or graphical buttons to select object features.

FIG. 5 illustrates, for example, graphical check boxes to select object features including size, shape, intensity, texture, location, area, perimeter, shape factor, equivalent diameter, length, width, integrated fluorescence intensity, mean fluorescence intensity, variance, skewness, kurtosis, minimum fluorescence intensity, maximum fluorescence intensity, geometric center, x-coordinate of a geometric center

or y-coordinate of a geometric center. FIG. 5 illustrates a set including some of the most commonly used object features used to measure objects in an image. However, the present invention is not limited to the object features listed in FIG. 5 and more, fewer or equivalent object features can also be used. FIG. 5 also illustrates graphical radio buttons for selecting fluorescence channels for desired dyes. Aggregate features are selected with a similar graphical user interface.

Returning to FIG. 4, at Step 68, a set of assay features is selected from multiple pre-determined assay features to analyze the set of images. In one embodiment of the present invention, Step 68 includes creating a protocol for an assay by selecting multiple pre-determined assay features (e.g., selecting multiple graphical buttons from FIG. 5). A "protocol" specifies a series of system settings including a type of analysis instrument, an assay, dyes used to measure biological markers, cell identification parameters and other general image processing parameters used to collect data. An "assay" is a specific selection of image processing methods used to analyze images and return results related to biological processes being examined. For more information on the image processing methods used in cell assays targeted to specific biological processes, see co-pending applications 09/031,217 and 09/352,171, assigned to the same Assignee as the present application, and incorporated herein by reference.

For example, for an exemplary assay-X, FIG. 5 illustrates selection of graphical check boxes for a perimeter 76, length 80 and width 82 object features for fluorescence channel zero, Dye-0 84. Radio button for DYE-0 84 is illustrated as selected in FIG. 5. Thus, assay-X would include obtaining object measurements for perimeters, lengths and widths of objects in images from fluorescence channel zero for a desired dye.

The assay features presented at Step 68 are associated with one or more image processing routines from a library of image processing routines to accomplish the assay feature measurement (e.g., at Step 60 of Method 52, FIG. 3). Thus, a user selecting the assay features presented at Step 68 does not have to understand how the assay feature is accomplished, but only how to choose desired assay features of interest to accomplish his/her own desired analysis (e.g., for a desired assay). If a new library of image processing routines was used, the assay features presented at Step 68 typically would not change, even though a whole new set of image processing routines might be used to accomplish an assay feature measurement.

Returning to FIG. 4, at Step 70 processing of the set of images using the selected set of assay features is requested. In one embodiment of the present invention, Step 70 includes selecting a series of general image processing operations in addition to selecting object and/or aggregate features. The image processing operations are applied before receiving the results at Step 78. The image processing operations may include filtering, object segmentation or mask modification (See, FIG. 6).

In one embodiment of the present invention, processing of the set of images at Step 70 includes applying general image processing routines to an image acquired from experimental data in a pre-determined order using a set of desired assay features selected from a graphical user interface (e.g., FIG. 6). However, the present invention is not limited to such an embodiment. In such an embodiment, pre-determining the order of applying the general image processing routines relieves a user of another image processing detail when he/she is creating an assay or protocol. Assay features are presented on a graphical user interface (e.g., FIG. 6) in the order that they are processed. For example, before segmenting an image, it is usually important to filter

the image to improve the efficiency of the segmentation. The filters may smooth and sharpen an image. Providing a pre-determined order helps make the creation of an assay or protocol simpler than if a user had to also determine a processing order himself/herself. The pre-determined processing order may also help a user more easily compare his/her results between or among several different experiments.

In one embodiment of the present invention, processing the set of images at Step 70 with selected object and aggregate features may include both independent and dependent processing of fluorescence channels. "Independent processing" refers to the creation of "independent masks" for each of the fluorescence channels. As is known in the art, a "mask" is one or more binary values used to selectively screen out or let through certain bits in a data value. Masking is typically performed by using a logical operator (AND, OR, XOR, NOT) to combine the mask and the data value.

"Dependent processing" refers to the use of a mask from one channel to derive a mask for analysis in another channel. This "derived mask" may be a simple copy of the parent mask or further processing may be applied to the parent mask. Feature extraction in the second channel occurs based on the derived mask.

For example, an approach to analyzing the cytoplasm-to-nucleus translocation of a transcription factor in a cell can be performed using derived masks. First, labeled nuclei are used to establish a mask. Second, a Transcription Factor ("TF") channel is setup to use a derived mask. The TF channel is defined as dependent on the nucleus channel. This copies the nuclei mask to the TF channel. The mask can be applied directly to measure a mean nuclear intensity of the TF, which is proportional to the amount of TF in the nucleus. Next, the mask is dilated a number of times and the binary exclusive OR/XOR function applied to the pair of masks. This leads to a ring shaped derived mask positioned over the peri-nuclear cytoplasm. Analysis within this

mask provides an estimate of the amount of TF in the cytoplasm. By calculating the ratio of the mean intensity within the nuclear mask in the TF channel and the mean intensity within the cytoplasmic ring mask in the TF channel, a measure of the cytoplasm-to-nucleus translocation can be established.

In one embodiment of the present invention, at Step 70, images from selected fluorescent channels are typically processed through a series of general image processing operations before analysis. Such general image processing steps are used to remove noise and help improve feature interpretation. The general image processing steps may include filtering, segmentation, etc. as is discussed below.

Table 4 illustrates independent general image processing operations. However, other independent image processing operations can be used and the present invention is not limited to the independent image processing operations illustrated in Table 4.

Filtering – The ability to perform smoothing, noise reduction, or local contrast adjustment such as edge enhancement processing on the images as a preliminary step to segmentation, depending on the image quality and the task.

- Smoothing - The smoothing method is based on a uniform, low pass 3 X 3 kernel.
- Sharpening - The sharpening method is based on a common, high pass 3 X 3 kernel.

Segmentation – Segmentation allows separation of an image into separate objects.

- Separate Grey – This method can be applied to segment a grayscale image into objects. There is one input parameter for the method, which relates to the contrast of the input image. The output of this method is a binary image that is overlayed on a grey scale image to show the object division.
- Threshold (Fixed) – A single user specified threshold can be used for images with very stable backgrounds and relatively good SNR. This is an alternative to the Separate Grey operation. The output of this method is a binary mask.

- Threshold (Auto) - A histogram-based method where the minimum intensity between two peaks can be determined automatically and then optionally corrected before applying. The output of this method is a binary image.
- Threshold - Threshold is setup interactively via a slider or by typing in a threshold value. When using a fixed threshold, the threshold value will be applied throughout the scan. When using an auto threshold, the auto threshold is computed for the current image and the correction coefficient is determined to make it match the one set manually. This coefficient will be applied to every threshold value determined during the scan.
- Fill Holes - This method provides a means of filling holes in binary masks that may occur during segmentation.
- Remove Border Objects- This method removes objects that touch the border of the image. Masks that touch the border often represent objects that are only partly within the image. The features extracted from such objects may not be non-representative of a complete object.

Mask Modification – Masks from the segmentation process may be modified by multiple cycles of erosion and dilation. This is useful for smoothing the outlines of the masks as well as creating masks that may be impractical from just the segmentation methods. The sequence of erode and dilate, or dilate and erode, helps to remove noise from a mask outline.

- Erode - Masks may be reduced in size by binary erosion for any number of cycles. Each erosion is a reduction in the size of the mask by removing perimeter pixels.
- Dilate -Masks may be expanded in size by binary dilation for any number of cycles. Each dilation ads an additional outline of 1 pixel in width.
- Remove Small- Small objects can be pieces of debris or they may form due to the segmentation operations. These objects may be removed. The size value is related to half of the width. It is the number of erosions needed to erase the object.
- Separate Binary- Provides a means of separating binary object masks.

Table 4.

Table 5 illustrates general image processing operations that are useful to apply to a dependent mask. However, other image processing operations can be used and

the present invention is not limited to the image processing operations illustrated in Table 5.

Dependent Masks
• Erode – Masks may be reduced in size by binary erosion for any number of cycles.
• Dilate- Masks may be expanded in size by binary dilation for any number of cycles.
• XOR- Masks can be combined by application of the exclusive OR binary operation. Thus creating a ring around an original nuclear mask. The ring can be expanded or contracted relative to the original nuclear mask while the width of the ring stays unchanged.

Table 5.

FIG. 6 is a block diagram illustrating an exemplary graphical user interface 86 for selecting general image processing operations. These operations, illustrated in Tables 4 and 5, are selected by inputting a number in the graphical box displayed, or by checking a graphical check box. If a graphical box has a value of zero, or a graphical check box is not checked, the general image processing operation is not executed. For example, as is illustrated in FIG. 6, no filtering is requested. However, grey scale segmentation 88 is selected, a value of 50 is used for the grey scale threshold 90. In addition, an independent mask is selected for dilating the mask for 2 cycles 92, and the XOR operation 94 is selected for a dependent mask.

In one embodiment of the present invention, processing at Step 70 includes obtaining measurements for selected object and aggregate features. Table 6 illustrates one possible implementation of the object features from Table 1 using the independent masks from operations in Table 4. However, the present invention is not limited to this implementation and other implementations can also be used.

Object Feature (Independent Mask)	Description
Area	Number of pixels inside an object (mask).
Perimeter	Number of pixels in an outline.

Object Feature (Independent Mask)	Description
Equivalent Diameter	Diameter of the circle with circle area = Area.
Length, Width	Longest and shortest sides of a bounding box that fits an object the best (after rotating it 180 degrees).
Area	Length * Width
Shape	Perimeter^2 / 4π *Area (this feature is not simply a combination of Area and Perimeter).
Integrated Intensity	Sum of intensities within an object (mask).
Mean Intensity	Integrated Intensity / Area.
Variance	Variance of intensities within an object (mask).
Skewness	Third statistical moment for intensities within an object (mask).
Kurtosis	Fourth statistical moment for intensities within an object (mask).
Min Intensity	Minimum intensity within an object (mask).
Max Intensity	Maximum intensity within an object (mask).
Geometric Center X	X coordinate of a geometric center of an object (mask) within a field (image).
Geometric Center Y	Y coordinate of a geometric center of an object (mask) within a field (image).

Table 6.

The feature set for dependent or derived masks is more limiting than the set for independent masks. One reason for this is that dependent masks are not necessarily related to a form of a signal in a dependent channel. Thus, for example, a perimeter or shape of a derived mask is typically more related to a primary channel rather than the dependent channel.

Table 7 illustrates one implementation of object features for dependent masks created using the aggregate operations from Table 5.

Object Feature (Dependent mask)	Description
IntegrIntIndMask	Integrated intensity under independent mask applied to current channel.
AveIntIndMask	Average intensity under independent mask applied to current channel.
IntegrIntRingMask	Calculated only if XOR is selected: Integrated intensity under ring mask applied to current channel
AveIntRingMask	Calculated only if XOR is selected: Average intensity under ring mask applied to current channel
IndMask2RingRatio	Calculated only if XOR is selected: Ratio of average intensity under independent mask applied to current

Object Feature (Dependent mask)	Description
	channel to average intensity under ring mask applied to current channel

Table 7.

In one embodiment of the present invention for processing at Step 70, a primary mask is applied and desired object features are extracted, a derived mask is applied and aggregate features are extracted. In one embodiment of the present invention, object features represent cell data and aggregate features represent the well-level or microplate level data for a population of cells in a well. However the present invention is not limited to such an embodiment and aggregate features for other types experimental data can also be used.

In one embodiment of the present invention, object and aggregate features are calculated and constrained by settings of aggregate "feature gates." "Feature gates" are provided to define sub-set of an object population that will contribute to an object or aggregate feature set. The feature gates include selection of a range including a lower and upper limit on the range. For example a feature gate for the object feature area may be set with a lower limit of zero and an upper limit of 2000. Thus, only objects (e.g., cells) that have an area between zero and 2000 pixels will be included.

Returning to FIG. 4 At Step 72, results are received from the processing of the set of images using the selected set of assay features. In one embodiment of the present invention, the results are written to a local database associated with the analysis instruments 12, 14, 16 or client computers 18, 20, 22. In another embodiment of the present invention, the results may also be propagated to the shared database 24 and/or the store archive 30.

In one embodiment of the present invention, results may be displayed using one of three display options illustrated in Table 8. However, the present invention is

not limited to three display options and more or fewer display options can also be used.

Display Option	Description
Every Processing Step	Images will be redisplayed after every step in the processing sequence for each channel.
Final Labeled Field Mask	Images for labeled independent channels will be displayed after all processing steps.
Masked Field Image	Gray scale images will be displayed without a background.

Table 8.

In one very specific embodiment of the present invention, Method 64 can be used in an automatic manner. In such an embodiment, a protocol is created to automatically accomplish the steps of Method 64 and store results in a database for later analysis. Such a very specific embodiment may used in conjunction with a HTS system. When a desired experiment is completed, a protocol may be automatically initiated and used to automatically accomplish the steps of Method 64.

FIG. 7 is a block diagram illustrating an exemplary screen display 96 for graphically displaying information acquired from images processed using a desired set of assay features. However, the present invention is not limited to this screen display and other screen displays, and more or less information can also be displayed, and the information can be displayed in different formats.

The screen display 96 includes a portion of an image of interest 98 for an object (i.e., a cell) acquired from an image 100 including multiple objects (i.e., a population of cells). The screen display 96 includes object feature data 102 measured from the image of interest 98, and aggregate data 104 and 106 measured from image 100 and nine other images (not displayed). The object feature data 102 and the aggregate data 104 and 106 displayed includes object and aggregate features selected at Step 68 of Method 64 (FIG. 4).

The image of interest 98 includes a magnified image of an individual cell identified by 98' in the image 100 including multiple objects. Screen display 96 illustrates exemplary assay feature data only for well A-3 illustrated by the blacked well 108 in the graphical illustration of a microplate 110 including 1536 wells.

These methods and system described herein may allow experimental data from high-throughput data collection/analysis systems including images to be analyzed. The methods and system can be used for, but is not limited to analyzing cell image data and cell feature data collected from microplates including multiple wells or bio-chips including multiple micro-gels in which an experimental compound has been applied to a population of cells. If bio-chips are used, any references to microplates herein, can be replaced with bio-chips, and references to wells in a microplate can be replaced with micro-gels on a bio-chip and used with the methods and system described.

The methods and system help provide a general purpose assay development tool. The methods and system allow a biologist, other scientist, or lab technician not trained in image processing techniques to quickly and easily design protocols and assays to analyze images acquired from experimental data (e.g., cells). The methods and system may improve the identification, selection, validation and screening of new drug compounds that have been applied to populations of cells. The methods and system may also be used to provide new bioinformatic techniques to manipulate experimental data including multiple digital photographic images.

It should be understood that the programs, processes, methods and systems described herein are not related or limited to any particular type of computer or network system (hardware or software), unless indicated otherwise. Various types of

general purpose or specialized computer systems may be used with or perform operations in accordance with the teachings described herein.

In view of the wide variety of embodiments to which the principles of the present invention can be applied, it should be understood that the illustrated embodiments are exemplary only, and should not be taken as limiting the scope of the present invention.

For example, the steps of the flow diagrams may be taken in sequences other than those described, and more or fewer elements may be used in the block diagrams. While various elements of the preferred embodiments have been described as being implemented in software, in other embodiments in hardware or firmware implementations may alternatively be used, and vice-versa.

The claims should not be read as limited to the described order or elements unless stated to that effect. Therefore, all embodiments that come within the scope and spirit of the following claims and equivalents thereto are claimed as the invention.

WE CLAIM:

1. A method for presenting analysis features for experimental data on a computer system, comprising the steps of:

(a) presenting a plurality of pre-determined assay features for analyzing images acquired from experimental data, wherein an assay feature includes one or more pre-determined measurements for an object in an image acquired from the experimental data;

(b) receiving a set of desired assay features selected from the plurality of pre-determined assay features;

(c) selecting one or more image processing routines from a library of image processing routines for an assay feature from the set of desired assay features, wherein the one or more image processing routines are used to accomplish the selected assay feature;

(d) associating the selected one or more image processing routines with the assay feature; and

(e) repeating steps (c) and (d) for other assay features in the set of desired assay features.

2. A computer readable medium having stored therein instructions for causing a central processing unit to execute the method of Claim 1.

3. The method of Claim 1 wherein an assay feature includes one or more measurements for cells in an image acquired from cell experimental data.

4. The method of Claim 1 wherein the step of presenting a plurality of pre-determined assay features for analyzing images includes presenting a plurality of object features or a plurality of aggregate features for analyzing images acquired from experimental data.

5. The method of Claim 4 wherein the object features include size, shape, intensity, texture, location, area, perimeter, shape factor, equivalent diameter, length, width, integrated fluorescence intensity, mean fluorescence intensity, variance, skewness, kurtosis, minimum fluorescence intensity, maximum fluorescence intensity, geometric center, an X-coordinate of a geometric center or a Y-coordinate of a geometric center of a cell.

6. The method of Claim 4 wherein the aggregate features includes sizes, shapes, intensities, textures, locations, nucleus area, spot count, aggregate spot area, average spot area, minimum spot area, maximum spot area, aggregate spot intensity, average spot intensity, minimum spot intensity, maximum spot intensity, normalized average spot intensity, normalized spot count, number of nuclei, nucleus aggregate intensity dye area, dye aggregate intensity, nucleus intensity, cytoplasm intensity, difference between nucleus intensity and cytoplasm intensity, nucleus area, cell count, nucleus box-fill ration, nucleus perimeter squared area or nucleus height/width ratio for a population of cells.

7. The method of Claim 4 wherein the aggregate features further include mean size, mean shape, mean intensity, mean texture, locations of cells, number of cells, number of valid fields, standard deviation of nucleus area, mean spot count, standard deviation of spot count, mean aggregate spot area, standard deviation of aggregate spot area, mean average spot area, standard deviation of average spot area, mean nucleus area, mean nucleus aggregate intensity, standard deviation of nucleus intensity, mean dye area, standard deviation of dye area, mean dye aggregate intensity, standard deviation of aggregate dye intensity, mean of minimum spot area, standard deviation of minimum spot area, mean of maximum spot area, standard deviation of maximum spot area, mean aggregate spot intensity, standard deviation of aggregate spot intensity, mean average spot intensity, nuclei intensities, cytoplasm intensities, difference between nuclei intensities and cytoplasm intensities, nuclei areas, nuclei box-fill ratios, nuclei perimeter squared areas, nucleus height/width ratios, or cell counts for a population of cells.

8. The method of Claim 1 wherein the step of selecting one or more image processing routines from a library of image processing routines includes selecting one or more image processing routines from a library of image processing routines to measure size, shape, texture, location or intensity of an object.

9. The method of Claim 1 wherein the step of associating the selected one or more image processing routines with the assay feature includes associating the selected one or more image processing routines with a graphical entity on a graphical user interface, wherein the graphical entity includes an assay feature name.

10. The method of Claim 1 wherein the images include digital images of cells or components of cells.

11. A method for analyzing experimental data on a computer system, comprising the steps of:

acquiring a set of images from experimental data on an analysis device;

selecting a set of assay features from a plurality of presented assay features to analyze the set of images, wherein an assay feature includes one or more pre-determined measurements for an object in an image acquired from the experimental data, and wherein an assay feature is associated with one or more image processing routines from a library of image processing routines to accomplish the assay feature;

requesting processing of the set of images using the selected set of assay features; and

receiving results from the processing of the set of images using the selected set of assay features.

12. A computer readable medium having stored therein instructions for causing a central processing unit to execute the method of Claim 11.

13. The method of Claim 11 wherein the step of acquiring a set of images includes acquiring a set of images from a desired experiment as a desired experiment is being conducted or acquiring a set of images from a database after a desired experiment after has been conducted,

14. The method of Claim 11 wherein the step of acquiring a set of images includes acquiring a set of images of cells or cell components in a population of cells.

15. The method of Claim 11 wherein the step of selecting a plurality of pre-determined assay features for analyzing the set of images includes selecting a plurality of pre-determined assay features from graphical entities on a graphical user interface.

16. The method of Claim 11 wherein the step of selecting a set of assay features from a plurality of presented assay features to analyze the set of images includes selecting object assay features, aggregate assay features or general image processing assay operations.

17. The method of Claim 16 wherein the object features include size, shape, intensity, texture, location, area, perimeter, shape factor, equivalent diameter, length, width, integrated fluorescence intensity, mean fluorescence intensity, variance, skewness, kurtosis, minimum fluorescence intensity, maximum fluorescence intensity, geometric center, an X-coordinate of a geometric center or a Y-coordinate of a geometric center of a cell.

18. The method of Claim 16 wherein the aggregate features include sizes, shapes, intensities, textures, locations, nucleus area, spot count, aggregate spot area, average spot area, minimum spot area, maximum spot area, aggregate spot intensity, average spot intensity, minimum spot intensity, maximum spot intensity, normalized average spot intensity, normalized spot count, number of nuclei, nucleus aggregate intensity dye area, dye aggregate intensity, nucleus intensity, cytoplasm intensity,

difference between nucleus intensity and cytoplasm intensity, nucleus area, cell count, nucleus box-fill ration, nucleus perimeter squared area or nucleus height/width ratio for a population of cells.

19. The method of Claim 16 wherein the aggregate features further include mean size, mean shape, mean intensity, mean texture, locations of cells, number of cells, number of valid fields, standard deviation of nucleus area, mean spot count, standard deviation of spot count, mean aggregate spot area, standard deviation of aggregate spot area, mean average spot area, standard deviation of average spot area, mean nucleus area, mean nucleus aggregate intensity, standard deviation of nucleus intensity, mean dye area, standard deviation of dye area, mean dye aggregate intensity, standard deviation of aggregate dye intensity, mean of minimum spot area, standard deviation of minimum spot area, mean of maximum spot area, standard deviation of maximum spot area, mean aggregate spot intensity, standard deviation of aggregate spot intensity, mean average spot intensity, nuclei intensities, cytoplasm intensities, difference between nuclei intensities and cytoplasm intensities, nuclei areas, nuclei box-fill ratios, nuclei perimeter squared areas, nucleus height/width ratios, or cell counts for a population of cells.

20. The method of Claim 16 wherein the general image processing assay operations include filtering, segmentation or binary mask modification.

21. The method of Claim 11 wherein the step of requesting processing of the set of images using the selected set of assay features includes requesting processing

the set of images using independent masks or dependent masks corresponding to individual assay features in the selected set of assay features.

22. The method of Claim 21 wherein operations used to create the independent masks include masks for smoothing, sharpening, separate grey-levels, grey level thresholds, filling holes, removing border objects, eroding, dilating, removing small objects or separating binary masks.

23. The method of Claim 21 wherein operations used to create the dependent masks include masks for eroding, dilating or performing an Exclusive OR operation on binary masks.

24. The method of Claim 11 wherein the step of requesting processing of the set of images using the selected set of assay features includes requesting processing of the set of images first using a set of general image processing assay routines from the selected set of assay features and then requesting processing of any object feature or aggregate features in the selected set of assay features.

25. The method of Claim 11 wherein the step of requesting processing of the set of images using the selected set of assay features includes processing the set of images in an order corresponding to an order of the selected set of assay features.

26. The method of Claim 11 wherein the step of receiving results from the processing of the set of images using the selected set of image processing routines includes receiving the results by redisplaying an image from the set of images on a graphical user interface on the analysis device after every step in a processing sequence for the image.

27. The method of Claim 11 wherein the step of receiving results from the processing of the set of images using the selected set of assay features includes receiving the results on a graphical user interface from a database associated with the analysis instrument.

28. The method of Claim 11 wherein the analysis device includes an analysis instrument or a client computer on a computer network.

29. A system for analyzing experimental data, comprising in combination:

a plurality of pre-determined assay features for analyzing a set of images acquired from experimental data, wherein an assay feature includes one or more measurements for an object in an image acquired from the experimental data;

a set of image processing routines from a library of image processing routines for accomplishing a selected assay feature, and associated with a selected assay feature;

a graphical user interface for presenting a set of assay features selected from the plurality of pre-determined assay features as graphical entities, and for presenting results of analyzing a set of images; and

an image analyzer for analyzing a set of images acquired from experimental data, wherein the image analyzer uses one or more of the set of image processing routines associated with an assay feature from a selected set of assay features to analyze the set of images, and for presenting results from analyzing a set of images on the graphical user interface.

30. The system of Claim 29 wherein the plurality of pre-determined assay features includes object features, aggregate features or general image processing features.

FIG. 1A

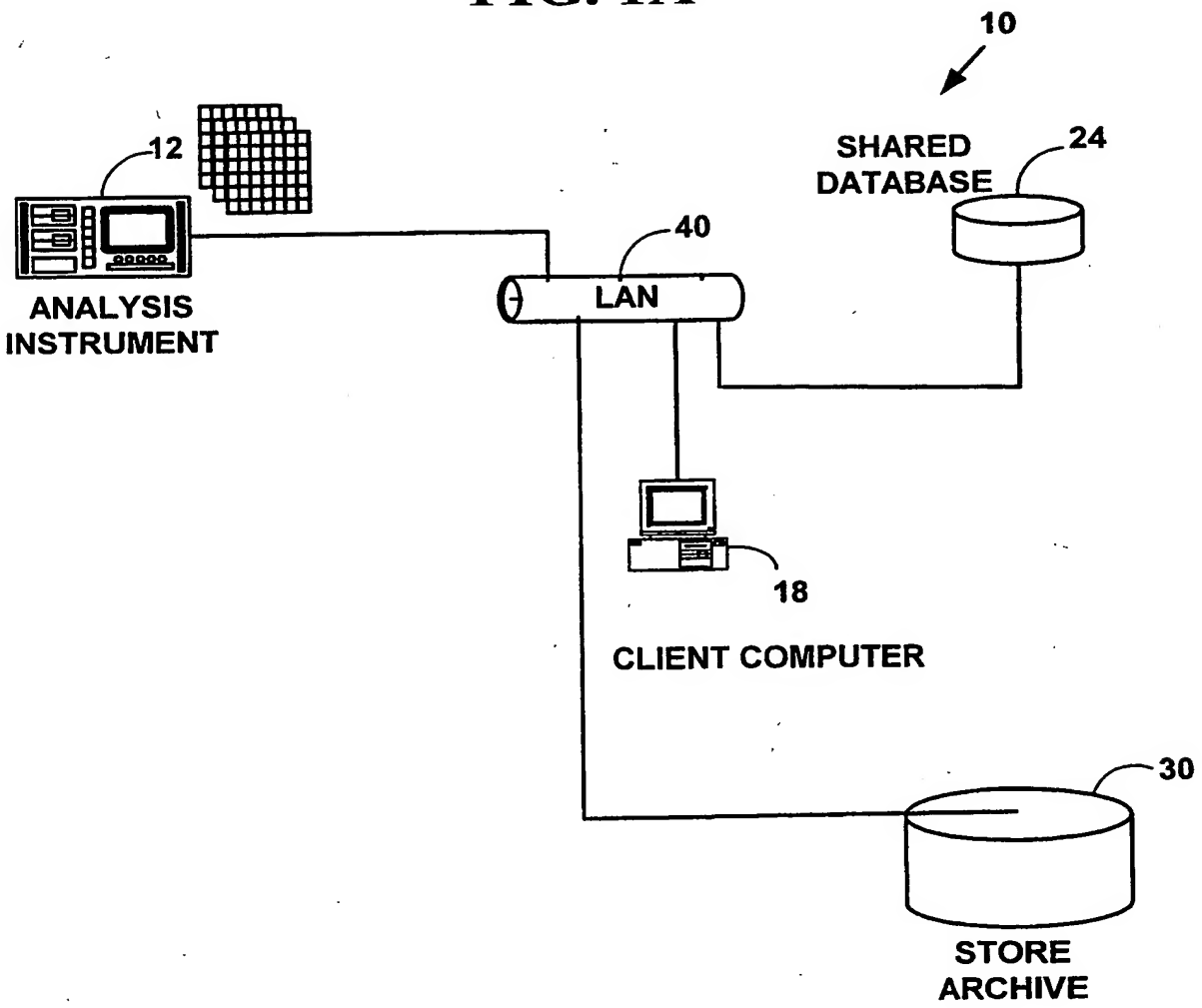

10
12
ANALYSIS INSTRUMENT
SHARED DATABASE
24
40
LAN
18
CLIENT COMPUTER
30
STORE ARCHIVE

FIG. 1B

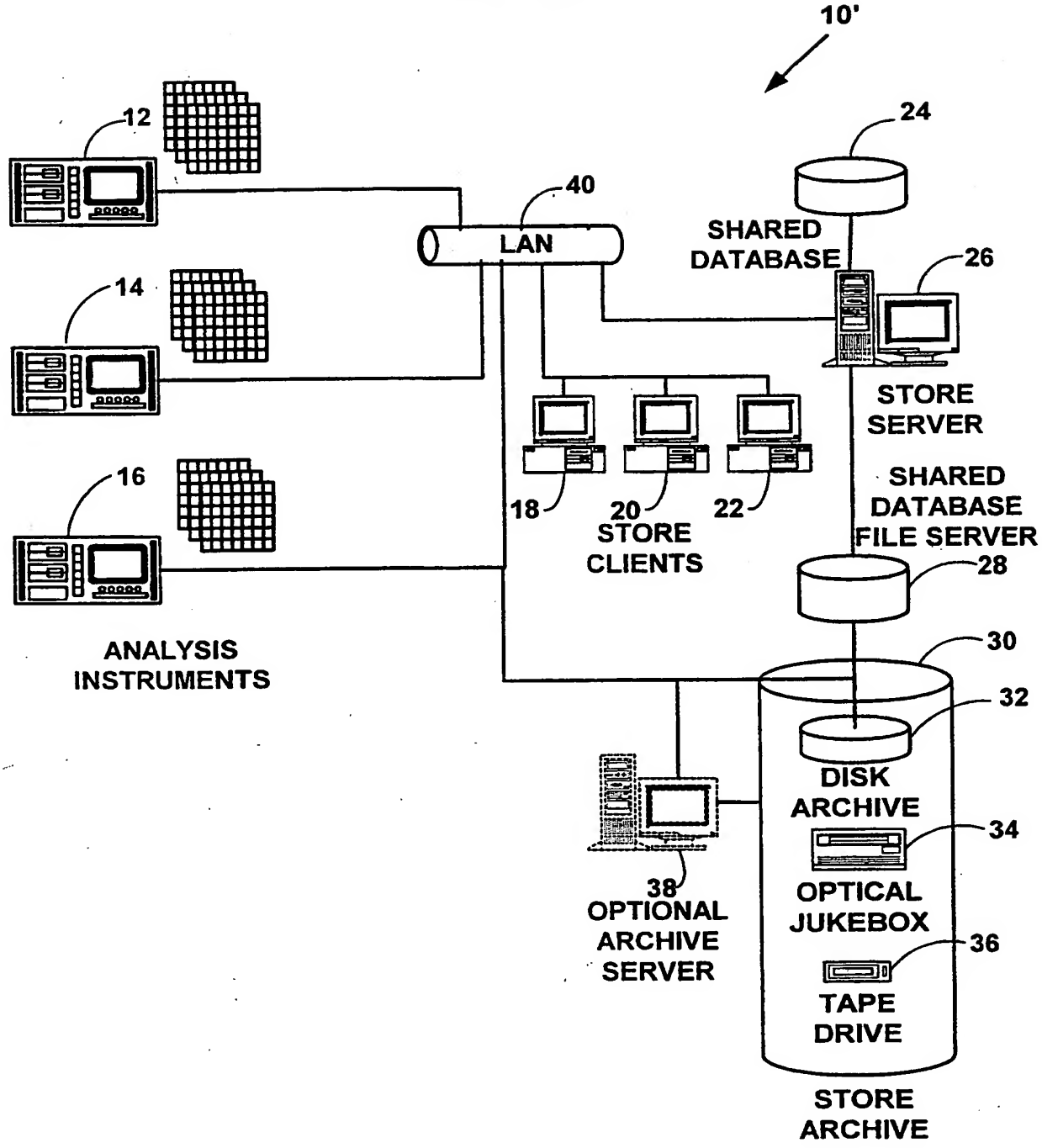

10'
12
14
16
ANALYSIS INSTRUMENTS
40
LAN
24
SHARED DATABASE
26
STORE SERVER
18
20
22
STORE CLIENTS
SHARED DATABASE FILE SERVER
28
30
32
DISK ARCHIVE
34
OPTICAL JUKEBOX
36
TAPE DRIVE
38
OPTIONAL ARCHIVE SERVER
STORE ARCHIVE

FIG. 2

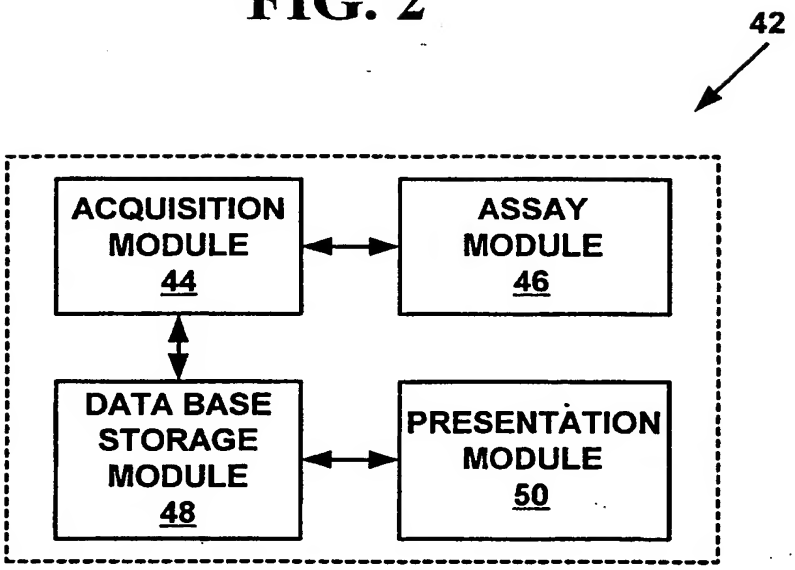

42
ACQUISITION MODULE 44
ASSAY MODULE 46
DATA BASE STORAGE MODULE 48
PRESENTATION MODULE 50
ARRAY SCAN MODULE

FIG. 3

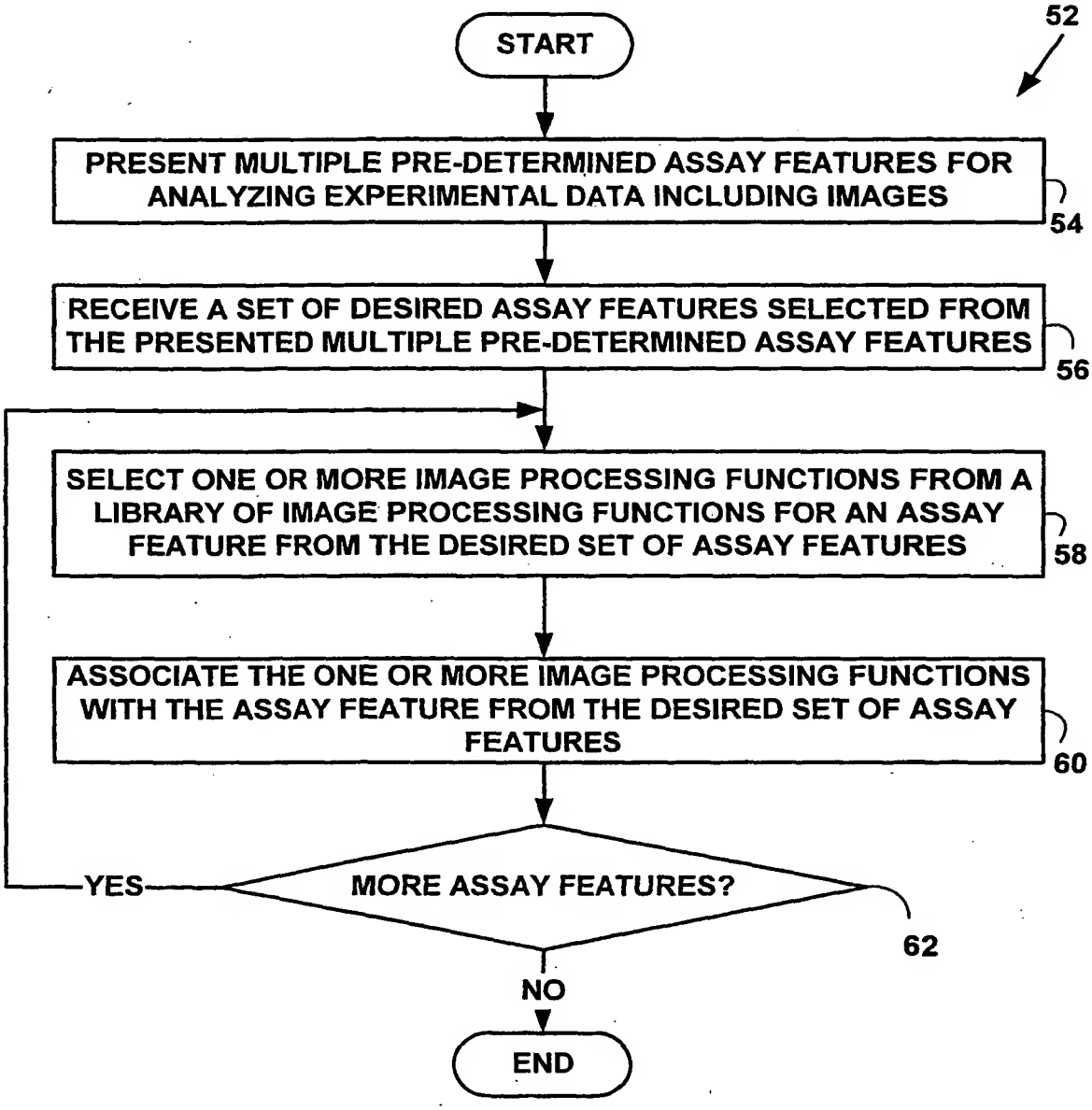

52
START
PRESENT MULTIPLE PRE-DETERMINED ASSAY FEATURES FOR ANALYZING EXPERIMENTAL DATA INCLUDING IMAGES
54
RECEIVE A SET OF DESIRED ASSAY FEATURES SELECTED FROM THE PRESENTED MULTIPLE PRE-DETERMINED ASSAY FEATURES
56
SELECT ONE OR MORE IMAGE PROCESSING FUNCTIONS FROM A LIBRARY OF IMAGE PROCESSING FUNCTIONS FOR AN ASSAY FEATURE FROM THE DESIRED SET OF ASSAY FEATURES
58
ASSOCIATE THE ONE OR MORE IMAGE PROCESSING FUNCTIONS WITH THE ASSAY FEATURE FROM THE DESIRED SET OF ASSAY FEATURES
60
MORE ASSAY FEATURES?
62
YES
NO
END

FIG. 4

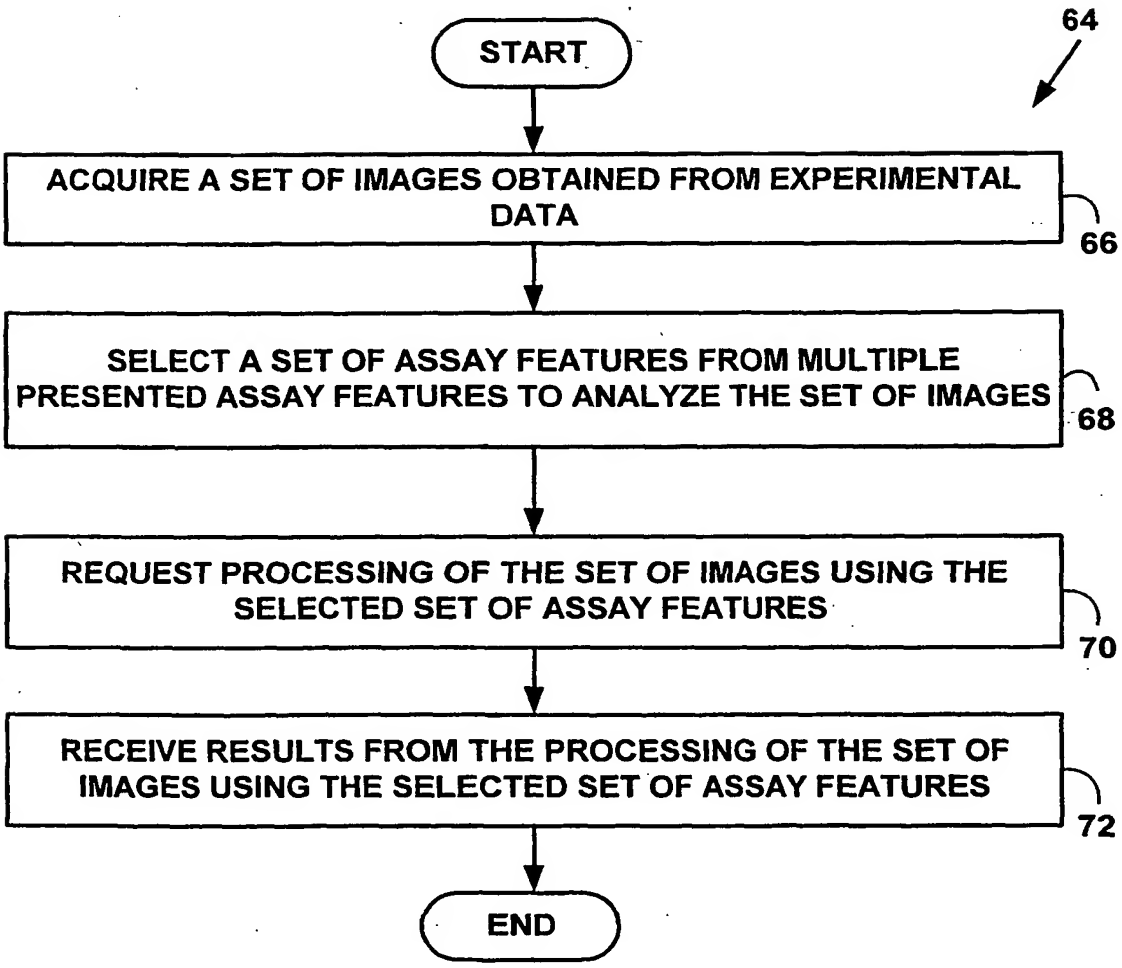

64
START
ACQUIRE A SET OF IMAGES OBTAINED FROM EXPERIMENTAL DATA
66
SELECT A SET OF ASSAY FEATURES FROM MULTIPLE PRESENTED ASSAY FEATURES TO ANALYZE THE SET OF IMAGES
68
REQUEST PROCESSING OF THE SET OF IMAGES USING THE SELECTED SET OF ASSAY FEATURES
70
RECEIVE RESULTS FROM THE PROCESSING OF THE SET OF IMAGES USING THE SELECTED SET OF ASSAY FEATURES
72
END

FIG. 5

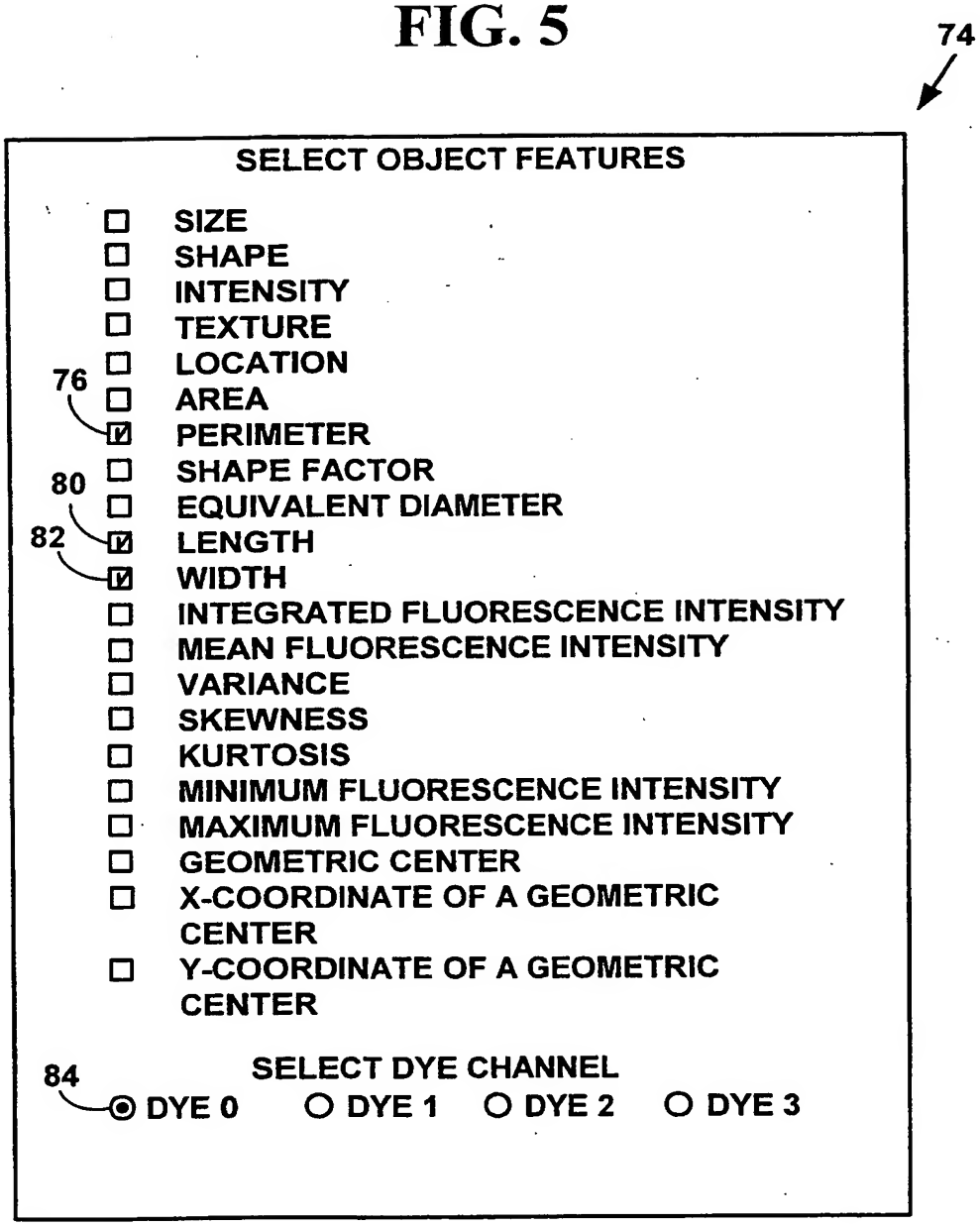

74
SELECT OBJECT FEATURES
SIZE
SHAPE
INTENSITY
TEXTURE
LOCATION
76
AREA
PERIMETER
SHAPE FACTOR
80
EQUIVALENT DIAMETER
82
LENGTH
WIDTH
INTEGRATED FLUORESCENCE INTENSITY
MEAN FLUORESCENCE INTENSITY
VARIANCE
SKEWNESS
KURTOSIS
MINIMUM FLUORESCENCE INTENSITY
MAXIMUM FLUORESCENCE INTENSITY
GEOMETRIC CENTER
X-COORDINATE OF A GEOMETRIC CENTER
Y-COORDINATE OF A GEOMETRIC CENTER
SELECT DYE CHANNEL
84
DYE 0
DYE 1
DYE 2
DYE 3

FIG. 6

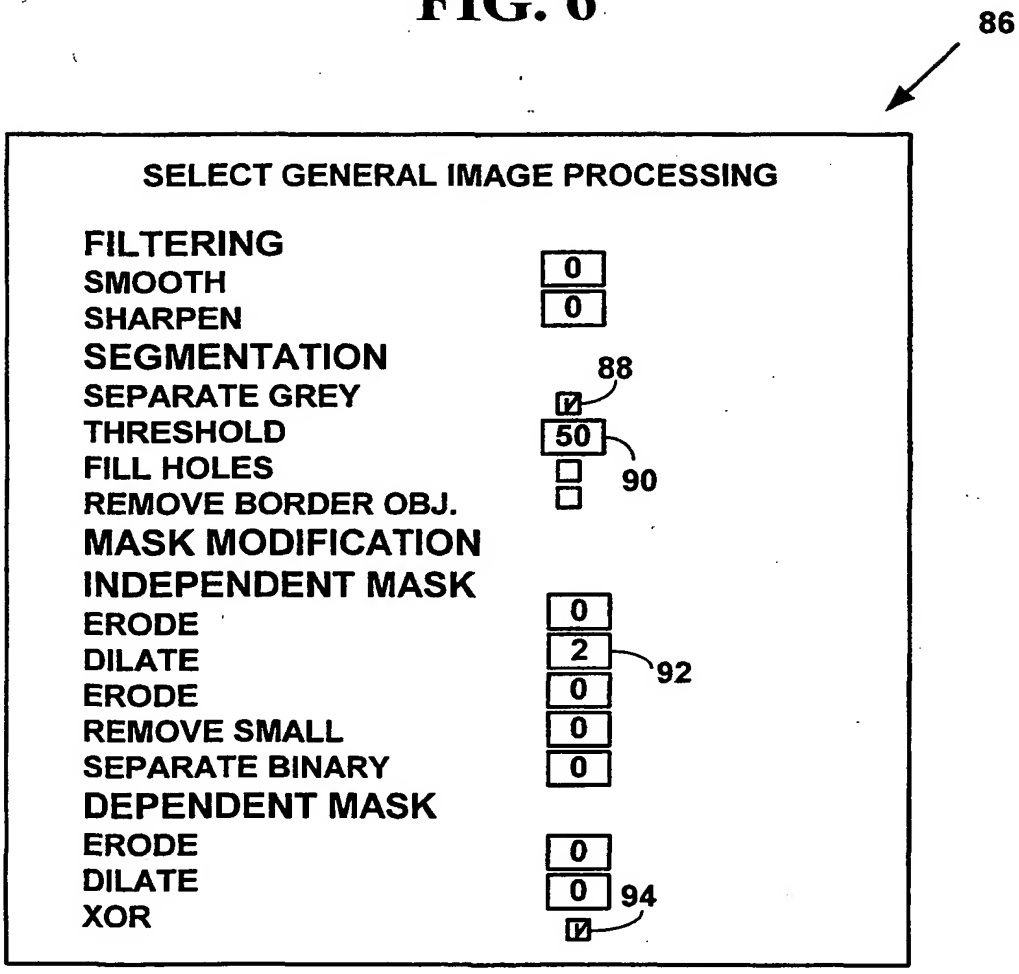
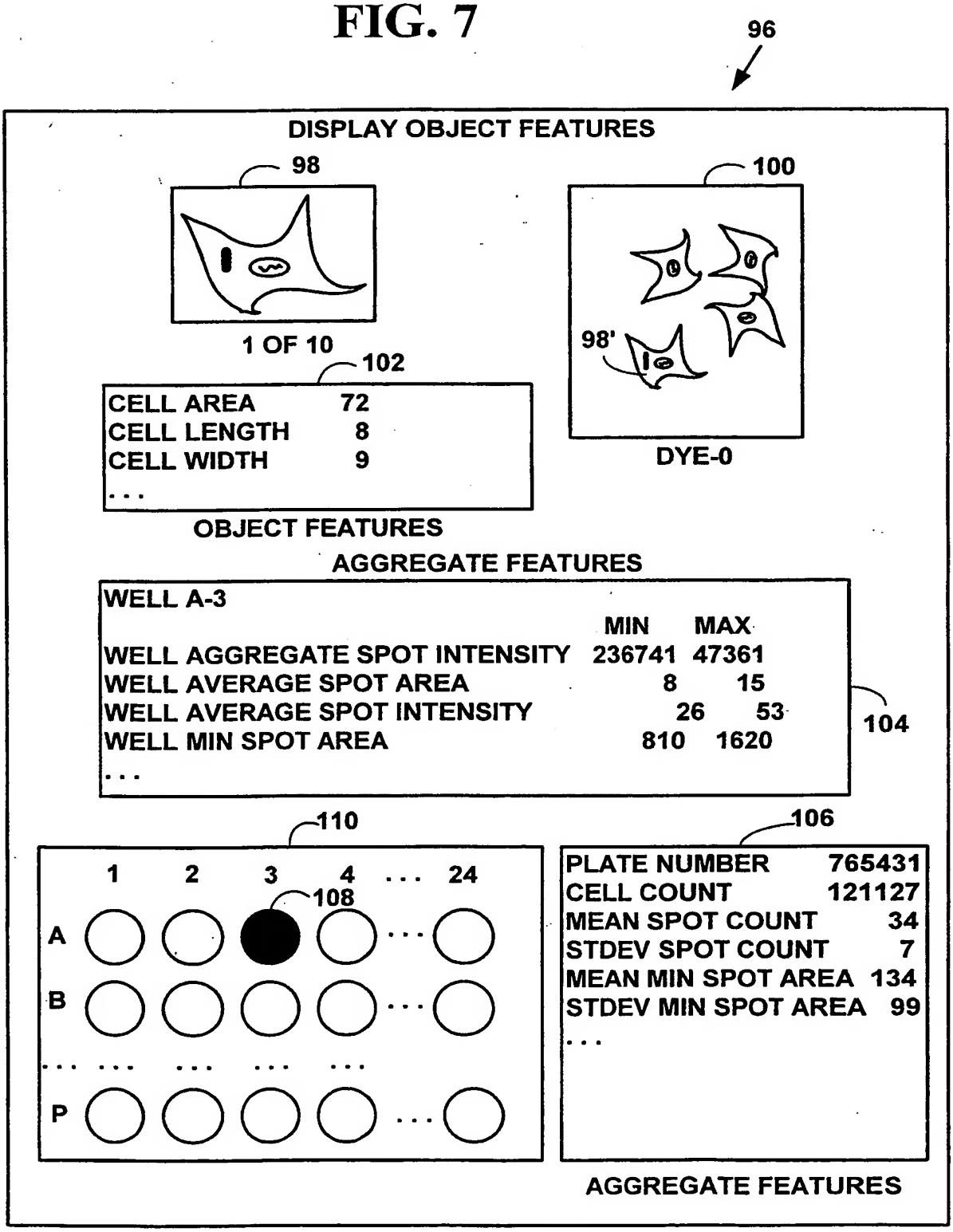

86
SELECT GENERAL IMAGE PROCESSING
FILTERING
SMOOTH
SHARPEN
SEGMENTATION
SEPARATE GREY
THRESHOLD
FILL HOLES
REMOVE BORDER OBJ.
MASK MODIFICATION
INDEPENDENT MASK
ERODE
DILATE
ERODE
REMOVE SMALL
SEPARATE BINARY
DEPENDENT MASK
ERODE
DILATE
XOR
0
0
88
50
90
0
2
92
0
0
0
0
0
94

FIG. 7


96
DISPLAY OBJECT FEATURES
98
100
1 OF 10
98'
DYE-0
102
CELL AREA 72
CELL LENGTH 8
CELL WIDTH 9
...
OBJECT FEATURES
AGGREGATE FEATURES
WELL A-3
MIN MAX
WELL AGGREGATE SPOT INTENSITY 236741 47361
WELL AVERAGE SPOT AREA 8 15
WELL AVERAGE SPOT INTENSITY 26 53
WELL MIN SPOT AREA 810 1620
...
104
110
106
1 2 3 4 ... 24
108
A
B
P
PLATE NUMBER 765431
CELL COUNT 121127
MEAN SPOT COUNT 34
STDEV SPOT COUNT 7
MEAN MIN SPOT AREA 134
STDEV MIN SPOT AREA 99
...
AGGREGATE FEATURES

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 November 2000 (30.11.2000)

PCT

(10) International Publication Number
WO 00/72258 A3

(51) International Patent Classification[7]: G06T 7/60

(21) International Application Number: PCT/US00/14246

(22) International Filing Date: 24 May 2000 (24.05.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/135,481	24 May 1999 (24.05.1999)	US
60/140,061	21 June 1999 (21.06.1999)	US

(71) Applicant *(for all designated States except US)*: CELLOMICS, INC. [US/US]; 635 William Pitt Way, Pittsburgh, PA 15328 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: GOUGH, Albert, H. [US/US]; 2014 Louise Drive, Glenshaw, PA 15116 (US). LAPETS, Oleg P. [RU/US]; 1937 Shady Oak Circle, Allison Park, PA 15101 (US). BRIGHT, Gary [US/US]; 586 Sturbridge Drive, Highland Heights, OH 44143 (US).

(74) Agent: HARPER, David, S.; McDonnell Boehnen Hulbert & Berghoff, Suite 3200, 300 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

(88) Date of publication of the international search rep rt:
31 May 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/72258 A3

(54) Title: METHOD AND SYSTEM FOR GENERAL PURPOSE ANALYSIS OF EXPERIMENTAL DATA

(57) Abstract: Methods and system for general purpose analysis of images acquired from experimental data collected with automated feature-rich, high-throughput experimental data collection systems. A set of pre-determined general assay features is presented. An assay feature includes one or more measurements for an object in a digital photographic image acquired from the experimental data. The set of pre-determined general assay features includes object features, aggregate features and general purpose image processing features. A set of desired assay features is selected from the set of features. A set of images is processed using the desired assay features from the selected set of assay features. The methods and system help provide a general purpose assay development tool. The methods and system allow a biologist, other scientist or lab technician not trained in image processing techniques to quickly and easily design protocols and assays to analyze images acquired from experimental data (e.g., cells). The methods and system may improve the identification, selection, validation and screening of new drug compounds that have been applied to populations of cells. The methods and system may also be used to provide new bioinformatic techniques to manipulate experimental data including multiple digital photographic images.

INTERNATIONAL SEARCH REPORT

Intern. al Application No
PCT/US 00/14246

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06T7/60

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 235 522 A (JAMES W. BACUS) 10 August 1993 (1993-08-10) column 8, line 45 - line 52 column 9, line 10 - line 32; figures 7-11	1-30

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 November 2000	21/11/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Chateau, J-P

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

...ormation on patent family members

Internat...nal Application No
PCT/US 00/14246

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5235522 A	10-08-1993	US 5526258 A	11-06-1996
		US 5546323 A	13-08-1996